DOCUMENT RESUME

ED 169 135 TM 008 749

AUTHOR Bessemer, David W.; Spencer, Mary L.

TITLE Functional Literacy in Schoolchildren. Definition and Criteria of Test Selection.

INSTITUTION System Development Corp., Santa Monica, Calif.

SPONS AGENCY Pacific Training and Technical Assistance Corp., Berkeley, Calif.
PUB DATE 29 Aug 75
NOTE 44p.; For related document, see TM 088 766

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Achievement Tests; Basic Skills; *Calculation; Compensatory Education Programs; *Criteria; *Functional Illiteracy; *Functional Reading; Intermediate Grades; Junior High Schools; *Literacy; Reading Achievement; *Test Selection; Test Validity
IDENTIFIERS Elementary Secondary Education Act Title I

ABSTRACT

As part of the development of a functional literacy test for fourth through eigth grade children in Title I compensatory education programs, this report defines functional literacy for children and enumerates criteria for evaluating existing tests. Criteria for selection of a test include: (1) content, empirical, and construct validity; (2) appropriateness of the test for the examinee population, including disadvantaged children; (3) technical quality, including reliability and normative standards; and (4) usability by non-expert personnel, objective machine scoring procedures, and interpretation. The test should be an independent entity rather than components of an achievement test battery. The test will be in English, but a Spanish-language version should also be available. (MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED169135

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

FUNCTIONAL LITERACY IN SCHOOLCHILDREN

DEFINITION AND CRITERIA OF TEST SELECTION

BY

David W. Bessemer
Mary L. Spencer

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Mary Spencer

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Submitted to: System Development Corporation
2500 Colorado Avenue
Santa Monica, California 90406

Submitted by: Pacific Consultants
3099 Telegraph Avenue
Berkeley, California 94705

August 29, 1975

TM008 749

FUNCTIONAL LITERACY IN SCHOOLCHILDREN

DEFINITION AND CRITERIA OF TEST SELECTION

I. INTRODUCTION

Pacific Consultants is under contract to Systems Development Corporation (SDC) to assist in the selection or development of a test of Functional Literacy for use in SDC's Title I Evaluation of Compensatory Education Programs. This report presents the results of the first task in Pacific Consultant's work -- the definition of functional literacy and criteria for evaluating existing tests.

II. GENERAL REQUIREMENTS

An overall description of the characteristics desired in the test of Functional Literacy was contained in SDC's Statement of Work for the Title I evaluation, and further clarified in meetings between SDC personnel and representatives of Pacific Consultants.

First of all, the test must clearly measure functional literacy. For the purpose of the Title I evaluation, functional literacy is viewed as the reading and computational skills needed by youth as they deal in the contemporary non-school related world. It must be an

independent test designed to measure functional literacy rather than the reading and computation portion of an achievement test battery. The level, range, and content of the test must be appropriate for elementary school children in grades 4-8, including children from disadvantaged backgrounds.

Costs of the test should be in the normal range of costs for comparable tests. The test must be capable of group administration by non-expert school personnel employing uniform procedures across the country. It should be amenable to objective machine scoring procedures.

If a norm-referenced test is used, norms appropriate to the population of the study should be available. The norms should include appropriate age levels, ethnic groups, socioeconomic levels, geographical locations, and language considerations. If a criterion referenced test is used, the criteria on which a test is developed should have a valid relation to compensatory education goals and objectives.

Evidence of reliability must be available, preferably a coefficient of at least .80 based on one of the generally acceptable procedures for determining reliability. The computation of reliability should be based on data from a sample comparable to the population of the study. The

test must validly measure performance included in the functional literacy goals of compensatory education programs. Evidence of content validity, and/or validity based on comparisons with relevant criteria should be available.

The test must be suitable for the range and type of populations in the study. The test must be sufficiently current to reflect concerns for the propriety of content for special groups. The needs and characteristics of participants in the compensatory education programs suggest that the pupil population is highly diverse in ethnic background. Many pupils are below expected grade level in reading and computation, and some are considerably below grade level. Some pupils speak very little English. Nevertheless, the test should measure functional literacy in English because the ability to function in American society requires the performance of many tasks based on materials with English content. However, a Spanish-language version of the test should be available.

Finally, the test should be effective in meeting the major purposes of the Title I study, which consist of program evaluation and comprehensive research concerning compensatory education. The test should be sensitive to the effects of major variables, and sufficiently grounded in theory to permit hypotheses to be generated and findings to be interpreted in a meaningful fashion.

III. ASPECTS OF FUNCTIONAL LITERACY

Social indices of literacy have usually been based on years of schooling, or grade equivalent scores derived from standardized reading achievement tests. Assessments of literacy based on such indices have proved defective in two major ways. First, they apparently have led to a serious underestimation of the problem of literacy in the United States, and second, they have failed to provide a meaningful description of the actual capabilities of the population in terms suitable as a basis for policy. The defects of such indices have been summarized elsewhere (Harman, 1970; Bormuth, 1973; Nafziger et. al., 1975) and need not be reiterated here at length.

In recent years, an increasing emphasis has been placed on determining the actual abilities of adults performing practical reading tasks involved in real-world life situations, particularly those having social and economic utility. Harris (1970, 1971) surveyed adult performance in filling out common application forms, and answering questions based on newspaper employment advertisements. The National Assessment of Educational Progress (1971) conducted a nation-wide survey of adult performance on a number of reading tasks based on practical materials. The Army has successfully devised measures of reading, listening, and computational capabilities

required in specific military occupations based on job-related written materials (Sticht, et. al., 1972).

Tests designed specifically to measure adult functional literacy have been developed recently. In the course of developing such tests, very substantial progress has been made in the conceptualization of functional literacy, and in the methodology of literacy assessment.

The work conducted by Education Testing Service under support of the Office of Education's Targeted Research and Development Reading Program (now in the National Institute of Education) has established a solid body of data on the reading activities of American adults (Murphy, 1973). A survey based on a national probability sample identified the materials commonly read during everyday general activities, the duration of reading, and the perceived importance of the reading activity. Based on the survey, a large number of performance tasks were developed. Each task was classified by type of material, and type of socioeconomic benefit (or function) provided by the tasks. An advisory panel evaluated the importance of the tasks, and a national survey of adult performance was conducted on 170 selected tasks. The data of the survey was later used to estimate the economic value of functional literacy skills (Murphy, 1975).

A domain-referenced test of functional literacy (R/EAL) was developed by Lichtman (1974). Through logical analysis

she identified nine common classes of reading materials representative of everyday life activities and selected one specific type of material within each class. A specific behavioral objective was written for each type of material, and a task analysis conducted to identify subordinate objectives. Items were then written to cover the terminal and subordinate tasks in each hierarchy, using realistic facsimiles of actual sample materials.

To date, no efforts to develop functional literacy measures for schoolchildren in grades 4-8 have come to light. However, the previous work on adult functional literacy clearly indicates several primary dimensions which must be considered in defining the concept for school-age children, and which must be operationalized in the process of test development. These dimensions include:

1. The area of everyday life activity in which a functional literacy task is encountered.
2. The frequency and duration of encounters with the task.
3. The type of material involved in the task.
4. The specific behaviors required by the task.
5. The type of social function served by the task.
6. The importance of the task, as perceived by the individual taking the test.

7. The degree of socioeconomic benefit derived from successful performance, measured in economic terms, or as perceived by significant authorities.

One further dimension of the functional literacy domain has been suggested by Bormuth (1973). This is the linguistic characterization of the corpus of language associated with the materials. No serious consideration has as yet been given to linguistic factors in functional literacy, aside from identifying the mother-tongue used (e.g., English).

IV. DEFINITION OF FUNCTIONAL LITERACY

No definition of functional literacy has yet been generally accepted. Bormuth has suggested an all-inclusive definition of literacy in the following terms:

> "In the broadest sense of the word, literacy is the ability to exhibit all of the behaviors a person needs in order to respond appropriately to all possible reading tasks."

Nafziger, et. al. (1975) point out that literacy, unlike reading, refers both to basic reading skills and to socially appropriate reading behavior. They go on to indicate that *functional* literacy implies reading for a purpose, and a purpose related in some way to social utility.

The U.S. Office of Education has been quoted by Nafziger et. al. (1975) as defining a literate person

in the following terms:

> "...one who has acquired the essential knowledge and skills in reading, writing, and computation required for effective functioning in society, and whose attainment makes it possible for him to develop new aptitudes and to participate actively in the life of his times."

None of these definitions offer very specific guidance to the operationalization of the functional literacy concept as a basis for development of an effective assessment device. For the purpose of the Title I evaluation of Compensatory Education Programs, a definition is offered here which lends itself to operationalization in terms of the dimensions of functional literacy identified in the previous section.

Functional literacy of school children in the 4-8 grade range is the capability of performing in a consistently successful manner those reading and computational tasks which:

1. are normally encountered in the course of everyday life activities by a majority of children in non-school settings.
2. are normally encountered repeatedly, or involve a substantial duration of activity.
3. involve commerce with particular types of materials commonly found in the environment of the child.

4. involve a specific corpus of non-technical language and symbolic representation associated with the materials.
5. Require specific observable behavior in relation to the material.
6. serve definable types of social functions.
7. are regarded as important by the child or established authority figures, or have demonstrable and non-trivial socioeconomic benefits.

This definition is not fully operational, but can easily become so given the specification of standards for qualifiers and the drawing up of a number of systems of classification covering the dimensions mentioned in the definition. Suggestions for the operationalization of each aspect of the definition are presented in the following section.

V. REQUIREMENTS FOR OPERATIONALIZATION

Lacking a survey of reading activities for children in the 4-8 grade range comparable to that conducted with adults by ETS, it is relatively difficult to fully operationalize any definition of functional literacy. Nevertheless, a considerable amount of progress toward this goal can be accomplished through logical analysis, the advice of experts,

Table I

Areas of Life Activity for Children

1. Personal Maintenance
   a. Food
   b. Clothing
   c. Health
2. Personal Relations
   a. Family
   b. Friends
   c. Relatives
   d. Pets
3. Institutional Relations
   a. School
   b. Church
   c. Club
   d. Police and Fire
4. Neighborhood Locations
   a. Home
   b. Shopping
   c. Postal
   d. School
   e. Medical
   f. Recreation
   g. Library
5. Home Maintenance
   a. Appliances
   b. Yard and Garden
   c. Furniture
   d. Radio, T.V., Stereo
6. Travel
   a. Bicycle
   b. Bus
   c. Car
7. Leisure
   a. Sports
   b. Games
   c. Toys
8. Communications Media
   a. Newspapers and Magazines
   b. Comics
   c. Television and Radio
   d. Books
   e. Movies
   f. Mail
   g. Telephone

and by capitalizing on certain aspects of previous work done with adults.

The first major point of the definition proposed here is that tasks should be "normally encountered in the course of everyday life activities by a majority of children in non-school settings." This aspect can be operationalized in either one or both of two ways:

1. Given a comprehensive list of everyday life activities, SDC's literacy panel can be asked to nominate representative tasks which in their best judgement are "normally encountered by a majority of children."
2. Based on a comprehensive list of everyday life activities, a survey can be taken on a small but diverse sample of children who would nominate tasks which they could remember having performed in each area of life.

As a basis for this step, a tentative list of life activities is presented in Table I. In connection with each area of life, several significant entities are listed which might serve as effective stimuli-arousing associations to a reading or computation task. For example, a child asked to remember what he has computed relative to mom or dad, might remember having calculated one of the parent's ages, given the dates of birth. Or, asked about reading

Table II

Types of Materials

1. Signs, Labels
2. Schedules, Tables
3. Maps, Diagrams
4. Categorized Listings
5. Directions, Instructions
6. Advertisements, Announcements
7. Forms
8. Personal Communications
9. Instruments, Controls
10. Technical Documents
11. Discourse, Narrative

in connection with a post office or mail box, the child might remember reading the mail pickup times, and such labels as "local," "out-of-town," "air mail," "A.M.," "P.M.," etc.

If a survey of children is performed, data can also be gathered on the frequency and duration of each task, which can then be used to form basic standards for the second point in the definition. Otherwise, the question of frequency and duration will have to be left to expert judgement.

The third point has to do with the specific materials associated with a given task. The SDC literacy panel or the surveyed children can be asked to identify the materials used in connection with each task. However, it is also useful to have available a systematic classification system of materials, both as a stimulus to memory, and as a refinement and explication of the third point in the definition. Such a classification system is presented in Table II. The table has been based largely on the classification of materials used by ETS (Murphy, 1973) and Lichtman (1974).

There is presently no basis for explicitly defining the corpus of language and symbolic representations involved in children's functional literacy. Study of

the messages conveyed by specific materials included in an extensive sampling of materials would be required to empirically establish such a definition. Some restrictions might be proposed based on consultation with experts in child language development. Clearly, a child should not be expected to read language and handle concepts that are not yet incorporated into his spoken language competence, or that lie outside his normal realm of experience. One such restriction, confining the corpus to non-technical language has been included in the definition. Language should be considered technical if it occurs largely within a narrow field of social activity, is usually known primarily by persons directly engaged in that activity, and if most children are unfamiliar with the word, expression, or syntactic construction in question.

In the area of computation, similar restrictions can be proposed based on the normal course of school instruction. Clearly, no advanced forms of computation or symbols which children have not had an opportunity to learn should be included.

At this point in time, it is also rather difficult to propose a system of behavior categories which can be expected to provide a valid delineation of the specific

behaviors involved in functional literacy tasks. To a large extent, the nature of the reading tasks seem implicit in, and inferrable from, the material associated with the tasks. For example, the typical behavior involved in using a telephone directory is describable in terms of a systematic search algorithm, given a specific entry name, proceeding through subgoals defined by a sequence of alphabetic cues, and eventuating in location and retrieval of a specific numeric code from the listings. The usual behaviors associated with filling out forms include reading and comprehending headings or questions which identify requested information, retrieving personal information from memory or available records, and writing the information in appropriate blanks or checking off appropriate alternatives.

A systematic task-skills analysis of a considerable number of such tasks is required to define a comprehensive set of component skills covering the domain of functional reading tasks. Lacking such an analysis, an informal list of behavior categories is offered in Table III, based on a cursory examination of the 170 examples of adult materials used by ETS in constructing test items (Murphy, 1973) and those listed by Lichtman (1974). This list certainly fails to include some forms of behavior

Table III

Preliminary Classification System for
Behavior Exhibited in Functional Literacy Tasks

Behavior	Material
Systematic Search	Categorical Listings, Advertisements
Choice Discrimination	Signs, Labels, Maps, Diagrams, Instruments, Controls
Preference Selection	Categorized Listings, Advertisements
Retrieval of Personal Data	Forms
Selection and Storage of Information	Personal Communication, Discourse, Narrative
Performing Sequence of Operations	Directions, Instructions, Maps, Diagrams
Contingency Identification	Signs, Labels, Directions, Instructions, Technical Documents
Accuracy Verification	Technical Documents, Forms
Comprehension of Information	Personal Communication, Maps, Diagrams

particularly important for children, and others only relevant for adults, but it is a useful basis for later revision. Also listed in the table are the kinds of material which are most usually found associated with each behavior category.

A somewhat different tack can be taken in specifying the computational behaviors. Real-world computational problems can be distinguished on the basis of the computational operations, the numerical and measurement content, and the means of computation. This provides the basis for a facet design of computational categories shown in Table IV. As shown, the child will perform computational tasks associated with real-world materials requiring *operations*, based on *signed numbers* representing physical *quantities*, expressed in *units of measurement* by appropriate *means*.

ETS identified eight categories of socioeconomic functions benefits as shown in Table V. These seem to overlap to some extent with the categories of life activity presented in Table I. Revisions of these categories to make them more appropriate for children may be based on recommendations of SDC's literacy panel or consultation with experts in child development , or based on survey responses from children.

Table IV

Facet Design for
Computational Categories

Operations	Signs	Numbers
Addition	Positive	Integers
Subtraction	Negative	Fractions
Multiplication		Decimals
Division		Percentages
Quantities	**Units**	**Means**
Length	English	Mental
Area	Metric	Paper & Pencils
Volume	Kitchen	
Weight		
Time		
Dry Measure		
Liquid Measure		
Speed		
Money		
Cost/Unit		
Angular		

Table V

Functions of Literacy Behavior

1. Economic
2. Occupational
3. Education/ Culture
4. Recreation
5. Health
6. Maintenance
7. Personal Relationships
8. Citizenship

Evaluation of the importance of tasks can also be based on judgements of the Literacy Panel or survey responses. It may also be useful to have teachers and parents examine the degree of benefit to be attributed to specific tasks, since they bring somewhat different but significant perspectives to this issue.

VI. SELECTION CRITERIA

Characteristics of the test, the nature of the examinees, and the purpose of testing are important factors in selecting a test of functional literacy for use in the Title I Evaluation. The criteria for test selection presented here are based very largely on the general guidelines provided by the American Psychological Association's *Standards for Educational and Psychological Tests,* and the criteria employed in the test evaluations at the Center for the Study of Evaluation (CSE), as presented in the document *CSE Elementary School Test Evaluations*, authored by Ralph Hoepfner and others. Additional criteria were suggested by Pacific Consultants' previous review of reading and literacy tests for the Right-to-Read Evaluation, and the recent examination of tests of adult functional literacy performed at the Northwest Regional Educational Laboratory under the direction of Dean Nafziger.

The criteria suggested by the sources indicated above provided a reasonably complete compilation of factors relevant to test selection, but were not concerned specifically with the measurement of functional literacy in grades 4-8 for the purpose of program

evaluation. A number of general recommendations were not suitable in meeting the special requirements of the Title I Evaluation, and were therefore modified as necessary. The proposed criteria are organized according to the four general areas of 1) measurement validity, 2) examinee appropriateness, 3) technical quality, and 4) administrative usability in correspondence with the four areas identified in the CSE test evaluation system.

Validity

The criteria in this area concern the nature of what is measured by the test. It is most important that the test be clearly and unambiguously a measure of *functional literacy*, if its role in the Title I Evaluation is to be served. Factors contributing to the credibility of a test as measuring functional literacy are considered in terms of content, empirical, and construct validity.

A. Content Validity - It is highly desireable that the test be representative of a defineable population of items and performances with specific reference to the domain of functional literacy. The bases of definition and the procedures of test construction contribute

to content validity in terms of the criteria outlined below.

1. Definition - The test should be specifically designed as a test of functional literacy. Disagreement on the validity of content will surely arise if the test was originally designed for some other purpose, and if no explicit basis exists for judging the relevance of items.
2. Material Domain - The stimulus materials should be representative of those commonly encountered in real-life reading and computational tasks. Confidence in the representativeness of materials would be increased if a population of such materials were defined, the composition of the population was described in terms of types or characteristics of materials and formed part of the definition of functional literacy used as the basis of test development.
3. Behavior Domain - The performance required in the items should be representative of the tasks and skills commonly required in real-life reading and computational performances.

Explicit classification and/or description of a domain of functional literacy behaviors is desireable as part of the definition used as a basis for test development.

4. Symbolic Domain - The language and other symbolic representations which form the communicative component of the materials should be representative of the symbolic content commonly encountered in real-life reading and computational tasks. Specification of the symbolic content in linguistic and mathematical terms can further strengthen and clarify the definition of functional literacy beyond the material and behavior specifications usually considered. Such specifications could be particularly helpful in defining levels or ranges of competence in relation to the domains of materials and tasks.

5. Socioeconomic Domain - The materials and tasks should be representative of the socio-economic functions commonly encountered in real-life reading and computational tasks. A classification or description of socio-economic functions and the benefits or values

of performance should be part of the definition of functional literacy used as a basis of test development, to help insure that functionally significant rather than trivial performances are represented.

6. Program Objectives - The materials and tasks of the functional literacy test should *not* be referenced to specific program objectives. Program-referencing would amount to prejudging the result of the evaluation in relation to functional literacy, in that it would inevitably bias the evaluation in favor of program goals and those programs which emphasized the defined objectives. The test is intended to provide an objective criterion by means of which the effectiveness of various programs can be judged in the area of functional literacy.

7. Item Construction - Procedures used in constructing the items should be explicitly defined and replicable. Items should not permit response on the basis of mere word-matching between stem and answer. Rigorous item construction procedures are required to ensure the relevance and representativeness

of the item pool in relation to the aspects specified in the definition of functional literacy, whether material, behavioral, symbolic, or socioeconomic criteria are included. Procedures which are entirely algorithmic would be most advantageous but are not within the usual state of the art at present. Other procedures are acceptable if the resulting items show close correspondence to the classification systems employed in defining functional literacy.

8. Item Selection - Procedures used in selecting items from a pool for inclusion in the final test should be empirically based, and yield evidence that the representativeness of the pool was maintained or increased in the item selection process.

9. Subtests - The test should incorporate subtests for reading and computational tasks, but which may be related to the same stimulus materials.

10. Criterial Objectives - The definition of functional literacy should be supplemented and operationalized by the specification of a set of criterial tasks referenced directly

to the characteristics of materials, behavior, symbolic content, and functions employed in the functional literacy definition. Such objectives would provide an important link between definition and items. Such objectives might be used in item construction and selection, or as a basis for empirical validation of items.

B. Empirical Validity - It is desireable that the test have been used in previous studies, thus providing empirical evidence relating the test scores meaningfully to other variables. Areas of concern in relation to empirical validity are outlined below.

1. Concurrent Relations - It is advantageous but not essential that the test has been correlated in previous studies with a wide variety of other measures taken at the same time. The number and quality of studies, the number of variables, and the diversity of variables all contribute to the evidence bearing on the meaning of a given literacy score.

2. Predictive Relations - It is advantageous but not essential that the test have been

correlated with measures taken at some later time. The number and quality of studies as well as the number and diversity of variables contribute to the evidence bearing on the question of what consequences flow from having attained a given literacy score.

3. Causality - It is advantageous, though not essential, that studies have been performed which relate the functional literacy test to important psychological, educational, or socioeconomic independent variables. Such evidence should be of assistance in the analysis and interpretation of the findings in the Title I Evaluation.

4. Nature of Relations - Empirical relationships found in the available literature should be reasonably interpretable in terms of prevailing educational, psychological, and socioeconomic theory. The measure of functional literacy should relate sensibly to variables which can be expected to influence functional literacy, variables which can be considered to reflect components of functional literacy, and variables which are thought to be independent of functional

literacy. Factor analytic studies, if any are available, should indicate that the measure of functional literacy is factorially complex. The nature of one particular relationship is especially important. The subtests should not correlate too highly with standardized tests of reading ability or pure computational skills. Very high correlations of this sort would indicate that the test did not adequately represent the diversity of skills required in a functional literacy measure.

5. Sensitivity - It is advantageous that the magnitude of effects observed was substantial when the test was used as a dependent variable in experiments or evaluations. That is, the test should be sensitive to the effects of appropriate independent variables, so that there is some assurance that appropriate effects will be revealed in the Title I Evaluation as well.

C. Construct Validity - Criteria in this area have to do with the theoretical basis of the functional literacy concept. They are of lesser importance

in judging validity than content and empirical criteria given the practical concerns of the Title I Evaluation, but are valuable characteristics nonetheless.

1. Process Constructs - The conceptualization, development, and empirical validation of the test should be grounded on relevant psychological, linquistic, educational theory in the area of reading and computation. Particularly important in this respect is the availability of a task-skills analysis which would define the components of functional literacy, indicate hierarchical relations among components, and tie performance to basic cognitive information processing operations. Such a theoretical foundation is useful in generating hypotheses and interpreting results.

2. Acquisition Constructs - The conceptualization, development, and empirical validation of the test should be grounded in relevant psychological, linquistic, and educational theory in the areas of instruction and cognitive and language development. Such formulations would provide a basis for tying changes

in functional literacy to specific educational practices, and related developmental changes.

3. Socioeconomic Constructs - The conceptualization, development and validation of the test should be grounded on relevant social and economic theory to provide a basis for hypothesis and interpretations of findings concerning relevant socioeconomic variables, and the function and benefits of literacy.

Appropriateness

The second set of criteria concern the appropriateness of the test in relation to characteristics of the intended sample of examinees. The criteria focus on the three areas of instructions, items, and format and procedure.

A. Instructions

1. Clarity - The instructions should be appropriate in orientation and tone, inoffensive in content, and comprehensible with vocabulary and syntax suitable for children in the 4-8 grade range.

2. Purpose - The instructions should provide an honest explanation of its purpose and intended use.

3. Comprehensiveness - The instructions should precisely and completely describe all requirements of the tasks presented in the items so that the examinee has all the information needed to adopt an effective performance strategy.

4. Sample Items - The instructions should include sample items accurately illustrating task requirements and the level of difficulty of the tasks.

5. Mode - The instructions should be presented in an oral mode. If administratively feasible, the instructions should be tape recorded, or if not, a standardized script should be available suitable for fluid oral reading by non-expert examiners.

B. Items

1. Difficulty - Items should include a wide range of difficulties, including some items relatively easy for 4th grade children, and some items relatively difficult for 8th graders.

The distribution of difficulties should be somewhat flatter than is usual for an achievement test so that the same test can be used over the whole 4-8 grade range. Thus, approximately equal numbers of items with moderate levels of difficulty at each grade level should be included.

2. Motivation - The items should be relevant, up-to-date, and interesting for children in the 4-8 grade range so as to arouse intrinsic motivation in task performance, without extensive exhortations being required to induce cooperation and effort.
3. Propriety - The content of the items should not involve any invasion of privacy, or any sexist, racist, or otherwise offensive aspects of content.

C. Format and Procedure

1. Physical Quality - The paper should be of good quality, the print bold and readable, and the illustrations clear and up-to-date. Reproduction of materials involved in common reading and computational tasks should present realistic facsimilies of the actual materials, preferably including full-color reproductions.

2. Layout - The test should be effectively arranged and cued to facilitate recognition of items as units, the perception of the relation between item stems, answers, and examinee response, and the progression of successive items and pages.
3. Timing - The test should be time limited but permit most examinees to attempt most items within the time allowed. Sectioning of the test, with timing instructions for each section may help to maintain appropriate pacing in the brief time alloted for this test. Items at all difficulty levels should be represented in each section.
4. Response Mode - The response should be marked directly on the test sheet, in a fashion permitting machine scoring. No separate answer sheet should be required.
5. Complexity - Each item should require one simple and direct response, with no multiple steps or complications other than those intrinsic in the task represented by the item. Several items might be used based on the same stimulus materials provided the relationship of each item to the stimulus is clear.

Technical Quality

Criteria addressed in this section pertain to the reliability and comparability of the test scores, and the quality of normative standards.

A. Reliability

1. Comparability - Alternate forms should be based on parallel items with comparable item statistics. The forms should correlate .80 or above at every grade level in the 4-8 range.
2. Stability - Test-retest correlations should be .80 or above over brief time intervals, i.e., one month or less. Reliability coefficients could be lower over longer intervals, particularly when instructional experiences have intervened having a substantial effect on the level of functional literacy performance. However, in the case where no shift in level of performance has occurred, the reliability should remain above .70 for intervals up to one year.
3. Internal Consistency - High internal consistency is not a *necessary* criterion for the functional literacy test. A test which

is highly homogeneous is likely to be unrepresentative of the full diversity of tasks which should be sampled in a functional literacy test. In particular, items involving reading should only be moderately related to computational items. The correlation between reading and computational subtests, if present in the test, should correlate below .70, and preferably below .50.

B. Normative Standards

1. Data Available - Although normative data is not essential in view of the large sample to be tested in the Title I Evaluation, and the emphasis on program comparison in the evaluation, it will still be helpful to have some prior normative data available as a basis for comparison.
2. Normative Sample - It is desireable that normative data be available for the 4-8 grade range, and for adults as well.
3. Representative - It is desireable that the sample be representative of racial, ethnic, sex, geographic and socioeconomic strata, rather than the result of incidental sampling.

4. Reporting - It is desireable that normative data be reported separately as well as combined over the racial, ethnic, geographical, and socioeconomic strata represented in the sample.

5. Item Statistics - It is useful if item statistics are reported both for the whole sample and broken down by strata. Item difficulties are the most important statistic, but item discrimination indices and inter-correlations are useful as well.

Usability

Criteria considered in this area pertain to the practical aspects of administration, scoring, and interpretation.

A. Administration

1. Personnel - Non-expert school personnel should be capable of administering the test with very little training. The services of a specialist or a testing expert, or extensive training should not be required.

2. Scheduling - The test should require no more than 30 minutes of testing time (preferably

20 minutes) at one occasion of testing. Tests taking longer than 30 minutes should be easily modifiable for shorter length with no more than normally expected loss of reliability.

3. Conditions - The test should be capable of administration in usual classroom settings, to group sizes in the normal range for intact classroom groups, and without the necessity of special equipment (other than a cassett tape player, if it is decided to use taped instructions).
4. Scoring - The test should be scored in an objective manner by machine. Machine scoring should be highly fail-safe and reliable, without complex error checking routines to proof the results.
5. Components - The test materials should be entirely of the paper-and-pencil test variety, with no special manipulanda, slides, or other unusual components.
6. Cost - Costs should be in the normal range of paper-and-pencil tests having good quality paper and printing, including color reproduction.

B. Interpretation

1. Manuals - A high quality test manual should be available, one which meets the appropriate APA standards for test manuals. A supplemental brochure describing the test and how to interpret its scores should also be available for relatively unsophisticated consumers of the results.
2. Meaning - The test scores should be highly meaningful and understandable in terms of specific performance by a nontechnical audience including the general public. It would be most meaningful if a hierarchy of performance levels could be devised, in which a person placed at one level could be described as capable of a specific list of tasks, and all tasks listed at lower levels. However, this may be an unrealistic goal.
3. Scales - The primary test scores should be directly understandable in absolute terms without the use of complex conversions or scaling. Forms of scaling or conversion to standardized scores may be used as a

supplement to the primary scores or for use by audiences with a higher level of technical background.

4. Implications - It is desireable that the implications of given test scores for educational practice or public policy be clear and relatively direct. However, what is actually required to meet this criterion is not entirely certain.

REFERENCES

Nafziger, D. H., Thompson, R. B., Hiscor, M. D., and Owen, T. R. Tests of functional adult literacy: an evaluation of currently available instruments. Final Report, June, 1975, U. S. Office of Education OPBE Contract No. 300-75-0098. Northwest Regional Educational Laboratory, Portland, Oregon.

Murphy, R. T. Adult functional reading study. Final Report, December, 1973, National Institute of Education Grant No. OEC-0-70-4791(508). Educational Testing Service, Princeton, New Jersey.

Murphy, R. T. Adult functional reading study. Supplement, January, 1975. National Institute of Education Grant No. OEC-0-70-4791(508). Educational Testing Service, Princeton, New Jersey.

Lichtman, M. The development and validation of R/EAL, an instrument to assess functional literacy. Journal of Reading Behavior, 1974, 6, (2), 167-182.

Bormuth, J. R. Reading literacy: its definition and assessment. Reading Research Quarterly, 1973, 9, (1), 7-66.

Hoepfner, R., Bastone, M., Ogilvie, V. N., Hunter, R., Sparta, S., Grothe, C. R., Sharri, E., Hufano, L., Goldstein, E., Williams, R. S., and Smith, K. O. C CSE Elementary School Test Evaluation. UCLA Graduate School of Education, Center for the Study of Evaluation, Los Angeles, California, 1975.

Sticht, T. G., Caylor, J. S., Kern, R. P., and Fox, L. E. Project REALISTIC: determination of adult functional literacy skill levels. Reading Research Quarterly, 1972, 7, (3), 424-465.

Harman, D. Illiteracy: an overview. Harvard Educational Review, 1970, 40, (2), 226-243.

National Assessment of Educational Progress, Reading and Literature: General Information Yearbook. Report # 02-GIV, Washington, D.C.: U. S. Government Printing Office, 1971.

National Assessment of Educational Progress, Reading, Released Exercises, Report # 02-R-20, Washington, D.C.: U. S. Government Printing Office, 1971.

Harris, Louis and Associates, Inc. Survival literacy study. National Reading Council, September, 1970.

Harris, Louis and Associates, Inc. The 1971 National Reading Difficulty Index: a study of functional reading ability in the United States. National Reading Council, August, 1971.